2007 Economic Census Methodology
All Sectors (Sector 00)

Sources of the Data

For all sectors in these data files, large- and medium-size firms, plus all firms known to operate more than one establishment, were sent report forms to be completed for each of their establishments and returned to the Census Bureau. For most very small firms, data from existing administrative records of other federal agencies were used instead. These records provide basic information on location, kind of business, sales, payroll, number of employees, and legal form of organization.

Firms in the 2007 Economic Census are divided into those sent report forms and those not sent report forms. The coverage of and the method of obtaining census information from each are described below:

1. Establishments sent a report form:
    a. Large employers, i.e., all multiestablishment firms, and all employer firms with payroll above a specified cutoff. (The term "employers" refers to firms with one or more paid employees at any time during 2007 as shown in the active administrative records of other federal agencies.)
    b. A sample of small employers, i.e., single-establishment firms with payroll below a specified cutoff in industries for which specialized data precludes reliance solely on administrative records sources. The sample was stratified by industry and geography.
    c. Sample frame establishments (Construction sector only). The sample frame consisted of the entire construction universe; there were no subpopulations that were explicitly removed from the sample frame. The sample frame was compiled from a list of all construction companies in the active records of the Internal Revenue Service (IRS) and the Social Security Administration (SSA) that are subject to the payment of Federal Insurance Contributions Act taxes. Under special arrangements, to safeguard their confidentiality, the U.S. Census Bureau obtains information on the location and classification of the companies, as well as their payroll and receipts data from these sources. Unfortunately, these sources do not provide establishment level information for companies with multiple locations. For multilocation companies, the establishment level information is directly obtained from the U.S. Census Bureau's Company Organization Survey. For single-location companies, the IRS-SSA information is generally sufficient for assigning the company to a specific six-digit NAICS industry code.

United States
Census
Bureau
Economic Statistics

2. Establishments not sent a report form:
   a. Small employers, i.e., single-establishment firms with payroll below a specified cutoff, not selected into the small employer sample. Although the payroll cutoff varies by kind of business, small employers not sent a report form generally include firms with less than 10 employees and represent about 10 percent of total sales of establishments covered in the census. Data on sales, payroll, and employment for these small employers were derived or estimated from administrative records of other federal agencies. Some of the firms were sent a shorter version of a report form to obtain industry classification information only.
   b. All nonemployers, i.e., all firms with no paid employees during 2007. Sales information for these firms was obtained from administrative records of other federal agencies. Although consisting of many firms, nonemployers account for less than 10 percent of total sales of all establishments covered in the census. Data for nonemployers are not included in these data files, but are released in the annual *Nonemployer Statistics* series.
   c. Nonsample frame establishments (Construction sector only). There were a limited number of establishments included in the business register who were completely unclassified at the time of the economic census — construction sampling operation. These establishments were mailed a general classification card in early 2008. A portion of these were ultimately determined to be in-scope of the economic census — construction. Since this determination was not made until after the sample selection operation had been completed; these establishments were treated as a supplement to the original universe and were sampled independently for inclusion in the derived estimates.

The report forms used to collect information for establishments in all sectors are available at help.econ.census.gov/econhelp/resources/.

A more detailed examination of census methodology is presented in the *History of the Economic Census* at www.census.gov/econ/www/history.html.

Industry Classification of Establishments

The classifications for all establishments are based on the *North American Industry Classification System, United States, 2007* manual. Changes between 2002 and 2007 are discussed in the text at the beginning of this report. Tables at www.census.gov/epcd/naics07/ identify all industries that changed between the 2002 North American Industry Classification System (NAICS) and 2007 NAICS.

The method of assigning classifications and the level of detail at which establishments were classified depends on whether a report form was obtained for the establishment.

United States Census Bureau

Economic Statistics

1. Establishments that returned a report form were classified on the basis of their self-designation, product sales or shipments, and responses to other industry-specific inquiries.
2. Establishments without a report form:
    a. Small employers not sent a form were, where possible, classified on the basis of the most current kind-of-business or industry classification available from one of the Census Bureau's current sample surveys or the 2007 Economic Census. Otherwise, the classification was obtained from administrative records of other federal agencies. If the census or administrative record classifications proved inadequate (none corresponded to a 2007 Economic Census classification in the detail required for employers), the firm was sent a brief inquiry requesting information necessary to assign a kind-of-business or industry code.
    b. Nonemployers were classified on the basis of information obtained from administrative records of other federal agencies.

Basis of Reporting

The economic census is conducted on an establishment basis. A company operating at more than one location is required to file a separate report for each store, factory, shop, or other location. Each establishment is assigned a separate industry classification based on its primary activity and not that of its parent company.

Description of the Sample Frame (Construction Sector Only)

The major objective of the sample design was to provide a sample that would provide reliable estimates at the state by industry level. For sample efficiency considerations, the establishments in the initial 2007 construction frame were partitioned into two components for developing estimates within the sample frame. The details of each are described below:

1. Probability-proportionate-to-size (pps) sample. There were three non-overlapping strata for sample selection. An independent sample was selected within each state by industry cell. The details of each stratum were defined as:
    o Stratum 1. This stratum was comprised of approximately 12,000 establishments associated with multilocation companies. The establishments of these multiunit companies were included in the construction sample with certainty.
    o Stratum 2. This stratum was comprised of approximately 145,000 single-location companies that could be classified into a valid 2007 NAICS industry. These cases accounted for approximately 75 percent of the payroll associated with single-location companies in construction. The industry code for most of these establishments was determined from the special classification card that was mailed in early 2007. This group was partitioned into state by NAICS (six-digit) cells and an independent sample selected from each cell. Within each cell, a probability-proportionate-to-size (pps) sampling strategy was used. Under this

United States
Census
Bureau

Economic Statistics

approach, the probability of selection for the sample for larger establishments is higher than for smaller establishments. There were approximately 80,000 establishments selected from this group.

- Stratum 3. This stratum was comprised of the remaining single-location companies. For these companies, we did not have an updated 2007 NAICS industry code. The most recent classification information available for these companies was their 2002 NAICS. Using this 2002 NAICS industry code, this stratum was partitioned into state by NAICS (four-digit) cells; and an independent sample selected from each cell. Again, probability-proportionate-to-size sampling methodology was utilized. There were approximately 30,000 establishments selected from this group.

Subsequent to the initial census mail-out, companies that initiated operations in 2007 were identified via administrative sources. To assure proper representation of the entire in-scope population, simple random samples of these new operations were selected and mailed separately.

2. Estimation and variances. Based on the response data, establishments were assigned to the appropriate NAICS (six-digit) industry. At each level of tabulation, unbiased estimates were derived by summing the weighted establishment data where the establishment sample weight was equal to the inverse of its probability of selection for the construction sample.

The resulting estimates were generated from one of many possible samples and are subject to sampling variability. Estimates of this sample variability were independently derived at all levels of aggregation. These sampling variances were then aggregated to the publication levels for the computation of the relative standard errors.

Reliability of Data

All data compiled in the economic census are subject to nonsampling errors. Nonsampling errors can be attributed to many sources during the development or execution of the census:

- inability to identify all cases in the actual universe;
- definition and classification difficulties;
- differences in the interpretation of questions;
- errors in recording or coding the data obtained; and
- other errors of collection, response, coverage, processing, and estimation for missing or misreported data.

United States
Census
Bureau

Economic Statistics

The accuracy of these tabulated data is determined by the joint effects of the various nonsampling errors or by the joint effects of sampling and nonsampling errors. No direct measurement of these effects has been obtained except for estimation for missing or misreported data; however, precautionary steps were taken in all phases of the collection, processing, and tabulation of the data in an effort to minimize the effects of nonsampling errors.

The Census Bureau obtains limited information extracted from administrative records of other federal agencies, such as gross receipts from federal income tax records and employment and payroll from payroll tax records. This information is used in conjunction with other information available to the Census Bureau to develop estimates for nonemployers, small employers, and other establishments for which responses were not received in time for publication.

The Census Bureau recommends that data users incorporate this information into their analyses, as nonsampling error and sampling error could impact the conclusions drawn from economic census data.

Data presented in this report for the Construction sector are subject to sampling errors, as well as nonsampling errors. The estimates developed from the sample can differ somewhat from the results of a survey covering all companies in the sample lists, but are otherwise conducted under essentially the same conditions as the actual sample survey. The estimates of the magnitude of the sampling errors (the difference between the estimates obtained and the results theoretically obtained from a comparable, complete-coverage survey) are provided by the standard errors of estimates shown in the data files.

The particular sample selected for the construction sector is one of many similar probability samples that, by chance, might have been selected under the same specifications. Each of the possible samples would yield somewhat different sets of results, and the standard errors are measures of the variation of all the possible sample estimates around the theoretically, comparable, complete-coverage values.

Estimates of the standard errors have been computed from the sample data. They are presented in the form of relative standard errors that are the standard errors divided by the estimated values to which they refer.

In conjunction with its associated estimate, the relative standard error may be used to define confidence intervals, or ranges that would include the comparable, complete-coverage value for specified percentages of all the possible samples.

The complete-coverage value would be included in the range:

- From one standard error below to one standard error above the derived estimate for about two-thirds of all possible samples.

United States **Census** Bureau

Economic Statistics

- From two standard errors below to two standard errors above the derived estimate for about 19 out of 20 of all possible samples.
- From three standard errors below to three standard errors above the derived estimate for nearly all samples.

An inference is that the comparable complete-survey result would fall within the indicated ranges and the relative frequencies shown. Those proportions, therefore, may be interpreted as defining the confidence that the estimates from a particular sample would differ from complete-coverage results by as much as one, two, or three standard errors, respectively.

For example, suppose an estimated total is shown at 50,000 with an associated relative standard error of 2 percent, that is, a standard error of 1,000 (2 percent of 50,000). There is approximately 67 percent confidence that the interval 49,000 to 51,000 includes the complete-coverage total, about 95 percent confidence that the interval 48,000 to 52,000 includes the complete-coverage total, and almost certain confidence that the interval 47,000 to 53,000 includes the complete-coverage total.

In addition to the sample errors, the estimates are subject to various response and operational errors: errors of collection; reporting; coding; transcription; imputation for nonresponse, etc. These operational errors also would occur if a complete canvass were to be conducted under the same conditions as the survey. Explicit measures of their effects generally are not available. However, it is believed that most of the important operational errors were detected and corrected during the U.S. Census Bureau's review of the data for reasonableness and consistency. The small operational errors usually remain. To some extent, they are compensating in the aggregated totals shown. When important operational errors were detected too late to correct the estimates, the data were suppressed or were specifically qualified in the data files.

As derived, the estimated standard errors included part of the effect of the operational errors. The total errors, which depend upon the joint effect of the sampling and operational errors, are usually of the order of size indicated by the standard error, or moderately higher. However, for particular estimates, the total error may considerably exceed the standard errors shown. Any figures shown in the tables of this publication having an associated standard error exceeding 75 percent may be combined with higher level totals, creating a broader aggregate, which then may be of acceptable reliability.

Treatment of Nonresponse

Census report forms included two different types of inquiries, "basic" and "industry-specific." Data for the basic inquiries, which include location, kind of business or operation, sales, payroll, and number of employees, were available from a combination of sources for all establishments. Data for industry-specific inquiries, tailored to the particular kinds of business or operation

United States™
Census
Bureau

Economic Statistics

covered by the report form, were available only from establishments responding to those inquiries.

Data for industry-specific inquiries in this sector were expanded in most cases to account for establishments that did not respond to the particular inquiry for which data are presented. Unless otherwise noted in specific reports, data for industry-specific inquiries were expanded in direct relationship to total sales of all establishments included in the category. In a few cases, expansion on the basis of sales was not appropriate, and another basic data item was used as the basis for expansion of reported data to account for nonrespondents.

All reports in which industry-specific data were expanded include a coverage indicator for each publication category, which shows the sales of establishments responding to the industry-specific inquiry as a percent of total sales for all establishments for which data are shown. For some inquiries, coverage is determined by the ratio of total payroll or employment of establishments responding to the inquiry to total payroll or employment of all establishments in the category.

Disclosure

In accordance with federal law governing census reports (Title 13 of the United States Code), no data are published that would disclose the operations of an individual establishment or business. However, the number of establishments in a kind-of-business or industry classification is not considered a disclosure; therefore, this information may be released even though other information is withheld. Techniques employed to limit disclosure are discussed at http://www.census.gov/econ/census07/www/methodology/disclosure.html.

United States
Census
Bureau

Economic Statistics